



DWUDZIESTE PIERWSZE
FORUM
TELEINFORMATYKI®

25 września 2015 roku
Sesja Młodych Mistrzów
www.forumti.pl

**Budowa hurtowni danych oraz analiza internetowych ofert pracy w Polsce
z wykorzystaniem technologii Big Data**

Kamila Krupa
Dawid Dacewicz
Adrian Kwiatkowski
Piotr Lemański
Damian Raniszewski

Uniwersytet Mikołaja Kopernika w Toruniu



Projekt pod patronatem
Urzędu Statystycznego
w Bydgoszczy

mgr Olga Lewandowska

<http://bydgoszcz.stat.gov.pl>

„Tempo, w jakim gromadzimy dane, znacznie redukuje naszą zdolność do ich poprawnej analizy.”

Profesor Patrick Wolfe
London's Big Data Institute

Big Data

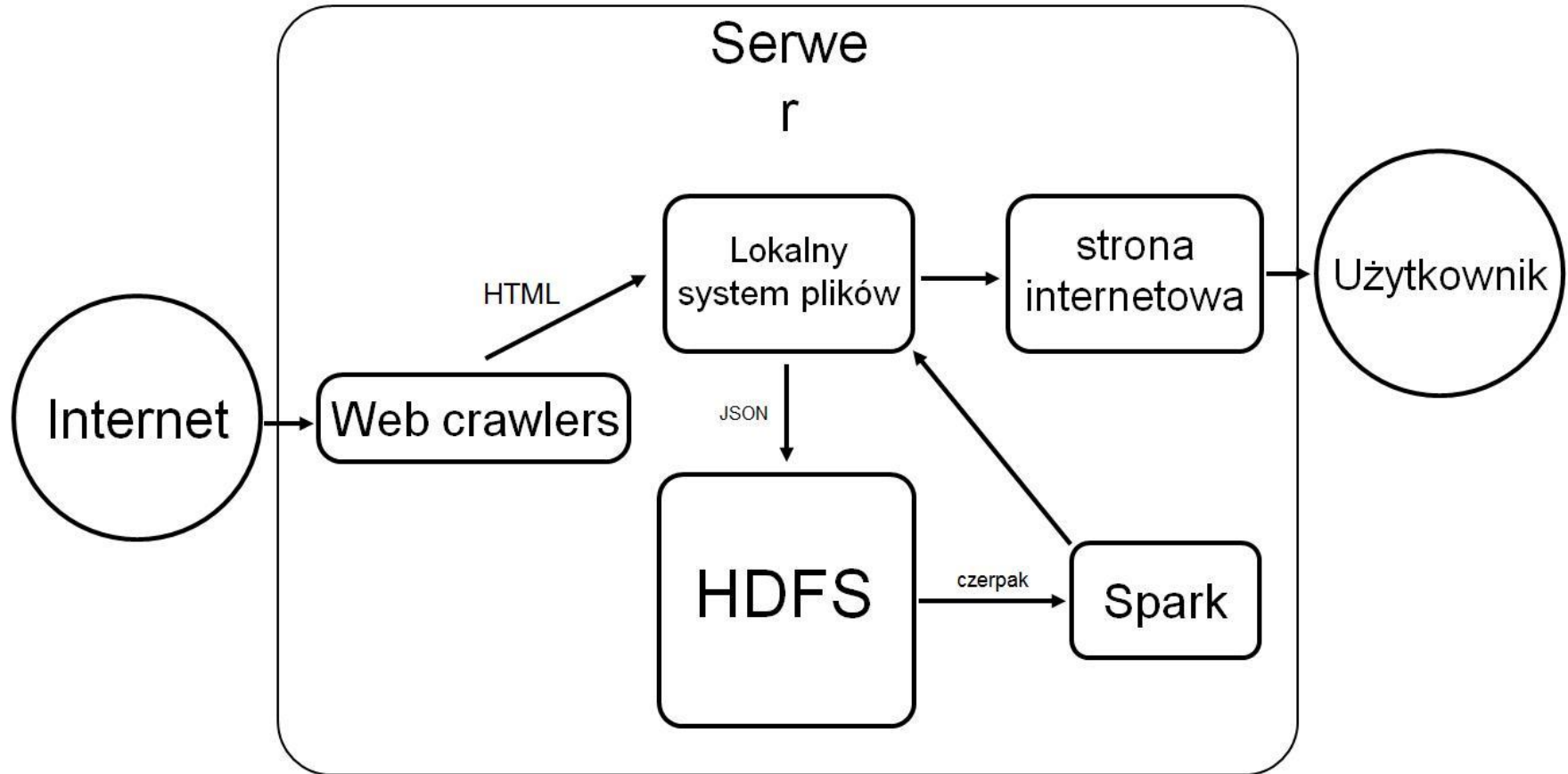
- Volume (ilość)
- Velocity (zmiennność)
- Variety (różnorodność)
- Value (weryfikacja)

Jak wygląda internetowy rynek pracy?
Mamy dużo danych, ale mało informacji.



Hurtownia danych umożliwiająca zdobycie informacji na temat internetowego rynku pracy w Polsce.

Struktura systemu



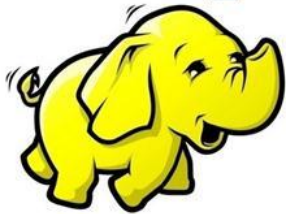
Wydajność

	Jeden węzeł	17 węzłów
Zasoby	2 GB RAM	68 GB RAM
	2 rdzenie	34 rdzenie
czas wykonywania skryptu	4 minuty 47 sekund	5 sekund

Technologie



hadoop



HTML



CSS



BeautifulSoup

JS



Dziękuję za uwagę